School of Computing and Information Systems
The University of Melbourne
COMP90049 Knowledge Technologies (Semester 1, 2019)
Workshop exercises: Week 8

1. Complete any remaining questions from the previous week, if necessary.
2. **Zoning** is the process of adjusting the weights of tokens, depending on document metadata (for example, whether the token appear in the page title, or an image caption, it might be assigned a different weight). Why is this a stage in the **parsing** step of an IR engine, and not the **indexing** or **querying** step?
3. What is a **phrase query**, and why is an inverted index — like the one from the question above — inadequate for phrase querying? How could the index be altered to support this style of querying?
4. What is **link analysis**? What aspects of user behaviour or the nature of data on the Web is it trying to model?
5. In the **PageRank** algorithm:
   (a) What is the mechanism for the "random walk"?
   (b) In terms of user behaviour, what is the significance of "teleporting"?
   (c) The lecture example of the PageRank algorithm was given as follows:

t	$\pi(d_{(1,t)})$	$\pi(d_{(2,t)})$
0	0.5	0.5
1	$0.5 \times 0.2 \times 0.5 + 0.5 \times 0.5 = 0.3$	$0.5 \times 0.2 \times 0.5 + 0.5 \times 0.8 + 0.5 \times 0.5 = 0.7$
2	$0.3 \times 0.2 \times 0.5 + 0.7 \times 0.5 = 0.38$	$0.3 \times 0.2 \times 0.5 + 0.3 \times 0.8 + 0.7 \times 0.5 = 0.62$
3	$0.38 \times 0.2 \times 0.5 + 0.62 \times 0.5 = 0.348$	$0.38 \times 0.2 \times 0.5 + 0.38 \times 0.8 + 0.62 \times 0.5 = 0.652$

   i. Draw the graph corresponding to this network (which is a little smaller than the World Wide Web!).
   ii. Which terms represent "following a link" and which represent "teleporting"?
   iii. What is the value of α in the above example? Re-do the above with $\alpha = 0.5$